https://doi.org/10.46335/IJIES.2021.6.10.8 **e-ISSN: 2456-3463**
Vol. 6, No. 10, 2021, PP. 41-44

International Journal of Innovations in Engineering and Science, www.ijies.net

Spam Detection in Online Social Network with K-Means Clustering and SVM Machine Learning Approach

Rucha Narkhede[1], Prof. Rahul Gaikwad [2]

[1]PG Student, Godavari College of Engineering, Jalgaon, Maharashtra, India
[2]Assistent Professor, Godavari College of Engineering, Jalgaon, Maharashtra, India

***Received on**: 16 July, 2021* ***Revised on**: 11 August, 2021* ***Published on**: 13 August, 2021*

Abstract- *Online social networks (OSNs) are becoming extremely popular among Internet users as they spend significant amount of time on popular social networking sites such as Facebook, Twitter and Google. These sites are turning out to be fundamentally pervasive and are developing a communication channel for billions of users. The dependence on these platforms for seeking opinions, news, and updates etc. is increasing. While it is true that OSNs have become a new medium for dissemination of information, at the same time, they are also fast becoming a playground for the spread of misinformation, propaganda, fake news, rumors and unsolicited messages. The huge amount of information available on these sites attracts the interest of cyber criminals who misuse these sites to exploit vulnerabilities for the illicit benefits like advertising some product or to attract victims to click on malicious links or infecting users system. Spam detection is one of the major problems now a day in social networking sites. Most previous techniques use different set of features to classify spam and non-spam users. In the paper, the machine learning algorithms such as K-Means Clustering, Support Vector Machine, and Emotional base spam word analyzer are applied and the spam words & emotions are detected from user's post.*

Keywords- Social Network, Spam detection, K-Means, Machine Learning, SVM

I - INTRODUCTION

Increase in popularity of social networking sites allows us to gather enormous amount of data and information about users. Large amount of data present on these sites attracts also malicious users. Such users use autonomous programs that act like human to steal the users personal information, spreading misinformation and propaganda. The Spammers, thereafter, successfully proliferates spam messages among their highly connected communities. Another way in which Spammers work is by sending the victim large number of direct messages called Direct Messaging (DM) spamming.
Online Social Network Vulnerabilities Large number of users and huge amount of information being shared increases security and privacy issues in online social networking sites (OSNs). User can face various type of attacks while using OSNs that include viruses which can be send by spammers to harm users system, personal information can be stolen by trust worthy third party to use for their own interests, Fake accounts can also harm user by gaining trust to effect users reputation. These bots send friend request randomly to the user selected from the list. If victim accepts their request then user start communicating with the victim friends which if accept request increase bots acceptance rate.

The spam message wastes the storage space of our personal computers, laptops, mobile, etc. Clicking an embedded malicious link can affect the user's system. [7] Research showed that users are likely to click on the links posted by unknown persons. Research gave complete details analysis on evasion tactics that were used by spammers. These analyses different techniques that are used by spammer to avoid detection also proposed different set

https://doi.org/10.46335/IJIES.2021.6.10.8 **e-ISSN: 2456-3463**
Vol. 6, No. 10, 2021, PP. 41-44

International Journal of Innovations in Engineering and Science, www.ijies.net

of features to classify spam users. After extracting features different machine learning classifier are used.

II - LITERATURE REVIEW

Online social networking sites are built on principles of trust and have attracted many researchers due to its popularity. They also purposed different features in messages that can be used for spam detection. These paper analyses different techniques that are used by spammer to avoid detection also proposed different set of features to classify spam users. After extracting features different machine learning classifier are used such as K-Means Clustering and Support Vector Machine techniques. [3]

Fabricio Benevenuto et al. detected spammers by identifying various user social behaviors and the characteristics of tweet content. These characteristics were used in a machine learning approach to classify the users as spammers and non-spammers. De Wang et al. in their study proposed a general framework to detect spam account across all the OSNs. The main contribution of their work was a new spam detected in any one social networking could be quickly identified across all other OSNs. Alex Hai Wang et al. proposed a model which uses a directed graph that depicts the relationship between "friends" and "follower" relationship in twitter. Bayesian Classifier was also used in his work, to detect spam accounts. Xin Jin et al. propounded a method for detecting spam accounts in social media network. They employed a GAD Clustering algorithm integrated with designed active learning algorithm to deal with spam messages. [2]

M. McCord and M. Chuah discussed various features related to user and tweet content which can be utilized in the detection of accounts intended for spamming. In their work, he evaluated four classifiers and compared there accuracies. Carlos Castillo et al. constructed a spam detection system that exploits the linked dependencies of web pages. The algorithm assumed that the linked web pages belonged to same class, i.e. if one web page is spam than its linked web pages must also be spam. Hongyu Gao et al. studied spam accounts in one of the popular OSN, Facebook. In their study, they found out that spamming was most common during early hours, when regular users were asleep. [1]

Authors presented a new application for Facebook users known as "MyPageKeeper" that detects spams. It protects users from spam attacks. MyPageKeeper uses content of profile instead of using other user information. SVM is used for detection of spammer from non spammers post using various features. MyPageKeeper considers post to be spam if some specific device is used for posting messages, post contain different false promises; post is created by using specific person without knowledge of that person. [5]

III -METHODOLOGY

3.1 Design Considerations

1. To classify the posts which is done by first taking out by split method involves first taking out whether the word is spam or not.
2. Rating the post after it is posted by detecting malicious keyword. The intensity of spam words shows whether the range of spam words are high, medium or low depends on red, orange and green zone.

3.2 Description of the Proposed System Approach

An overview of the complete process of spam detection is shown in the diagram in Figure 1, each of whose steps are explained in this section. The preliminary step for the detection of spammers in any OSN is data collection and necessary preprocessing to convert it into a form, which can be used by the learning algorithms.

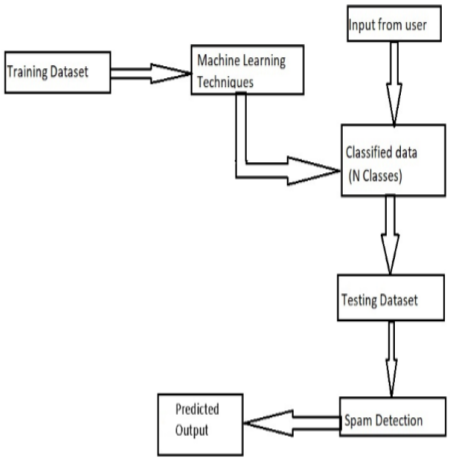

Fig. 1- Proposed Spam Detection Approach

Our proposed system approach is shown in Figure 1. At first user posts any message and it is split by split method. If the post is reported spam then system will identify the most common spam words in that post which will then be used for future reference. In the next part, when the user is posting post, spam words will be identified in real time and classified as depends on the intensity range that is stored in the database. Machine will restrict the user from posting only if there are more number of spam words and it is identified high level spam. In the news feed, post will be shown along with the level of spam.

https://doi.org/10.46335/IJIES.2021.6.10.8 **e-ISSN: 2456-3463**
Vol. 6, No. 10, 2021, PP. 41-44

International Journal of Innovations in Engineering and Science, www.ijies.net

Feature Identification Since, spammers behave differently from non-spammers; therefore we can identify some features or characteristics in which both these categories differ. Such hybrid features for spam detection are effective against evasion tactics. We aim to achieve higher accuracy by combing all these features. Various features which we have used to detect spam messages include:

Spam Words: We use most popular spam words and count the number of occurrence of these spam words in posts of users. As spammer uses these popular spam words to spread misinformation and to advertise their products, this feature can be vital to identify spam posts.

Replies: Since, information or message sent by a spammer is useless, therefore people rarely replies to its post. On the other hand, a spammer replies to a large number of posts in order to get noticed by many people. This pattern can be used in the detection of spammers.

Hash tags: Hash tags are the unique identifier ("#" followed by the identifier name) which is used to group similar tweets together under the same name. Spammers use large number of hash tags in their posts, so that their post is posted under all the hash tag categories and thereby gets wide viewership and is read by many.
Emotions: Emotions are the important feature which is used for when the user posts any message on social media, how user mood is is identified. Therefore it is easy to identify spam posts.

3.3 Preprocessor

Spam Words in the dataset, labeled as ID and Intensity, and were used for training the learning algorithms and also in accuracy calculations. In preprocessing step, all the continuous features were converted into discrete. The procedure adopted to select the intensity for a particular feature was obtained from according to which all spam words are arranged in order of their feature values. Processing begins from the first word in user posts, and it checks word by word whether the word is spam or not. [6]

IV - RESULTS

The system is to provide solution for detecting the malicious spam detection in online social networks using various machine learning algorithms. The system will stop spammer from posting malicious posts using machine learning algorithm based on keywords. In our work, we have used K-Means Clustering and Support Vector Machine as learning algorithms. Although, each of these approaches can be solely used to classify spam words, but in order to increase the accuracy, we have combined these approaches into an integrated algorithm in our work.

4.1 K-Means Clustering

K-Means Clustering is basically an unsupervised learning technique. User posts a message then the message is split by splitting method and the spam words are classified according to their intensity range. If the intensity of spam word is in between 0 to 30, then the spam words range is green that is low. If the intensity of spam word is in between 30 to 60, then the spam words range is orange that is medium. . If the intensity of spam word is in between 60 or above, then the spam words range is red that is high.

True Rate = (No. of spam messages truly classified / Total no. of messages) * 100% (1)

False Rate = (No.of spam messages falsely classified – True rate) * 100% (2)

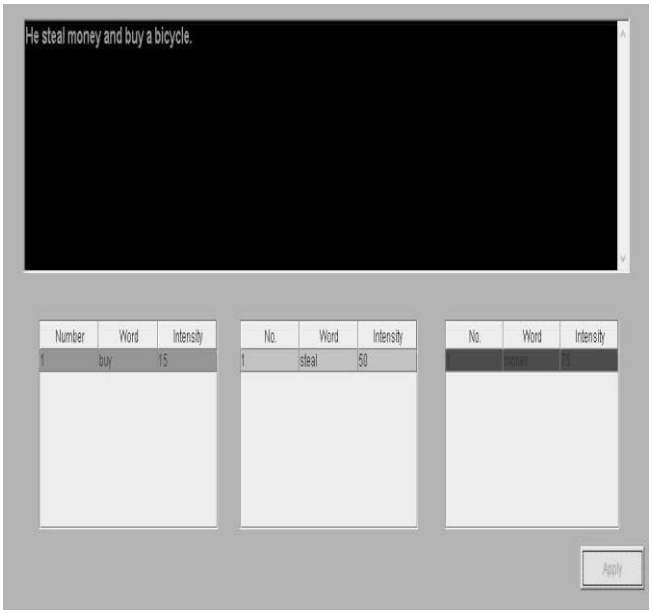

Fig.2- K-Means Clustering

4.2 Support Vector Machine (SVM)

SVM is one of the most powerful machine learning method. In this algorithm, graph shows the spam words are on the X-axis and intensity is on the Y-axis. After user posts a message the spam words will be identified according to their intensity. Greater spam words points are occur in which region, according to this the range of spam word is identified whether it is high, low or medium.

https://doi.org/10.46335/IJIES.2021.6.10.8 **e-ISSN: 2456-3463**
Vol. 6, No. 10, 2021, PP. 41-44

International Journal of Innovations in Engineering and Science, www.ijies.net

It can be concluded that using the spam and non-spam clusters based on the unsupervised K-Means Clustering is a effective method for detecting spam.

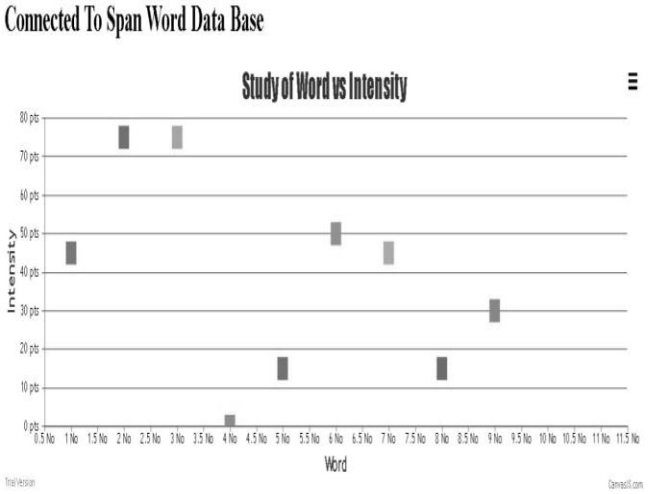

Fig. 3- Spam Word versus Intensity Graph

4.3 Emotional Base Spam Word Analyzer

In this algorithm, when the user post a message, user emotions are displayed whether user is sad, angry, happy, confused, love, etc. Spam words and emotions are stored in the database. After running the algorithm, one pie chart is displayed according to spam words in user posts and how is mood of user when he posts a message on social media. Percentages are also displayed based on emotions. [4].

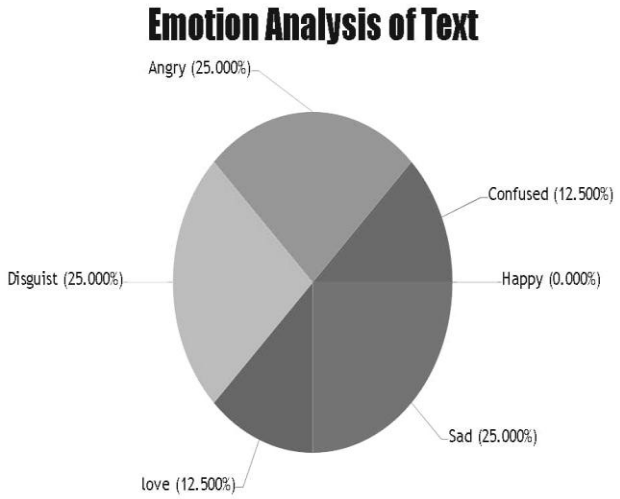

Fig. 4- Emotion Analysis of Text

V- CONCLUSION

Many machine learning techniques do exists that can be used for the spam detection in online social networks. In the paper, A systematic evaluation showing the promising performance of the spam detection framework with K-Means clustering and Support Vector Machine (SVM) learning techniques is presented. The method constructs spam detection model by the contents of various kinds of messages and find spam more efficiently. So it is concluded that using the spam and non-spam clusters based on the unsupervised K-Means Clustering is a effective method for detecting spam. In future, to enhance the system, other machine learning approaches can be used to detect spam images and fake accounts.

REFERENCES

[1] *Gupta, Arushi, and Rishabh Kaushal. "Improving spam detection in online social networks." In 2015 International conference on cognitive computing and information processing (CCIP), pp. 1-6. IEEE, 2015.*

[2] *Mateen, Malik, Muhammad Azhar Iqbal, Muhammad Aleem, and Muhammad Arshad Islam. "A hybrid approach for spam detection for Twitter." In 2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pp. 466-471. IEEE, 2017.*

[3] *Gheewala, Shivangi, and Rakesh Patel. "Machine learning based Twitter Spam account detection: a review." In 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), pp. 79-84. IEEE, 2018.*

[4] *Bhat, Sajid Yousuf, and Muhammad Abulaish. "Community-based features for identifying spammers in online social networks." In 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013), pp. 100-107. IEEE, 2013.*

[5] *Ponnapalli, HariGopal, Dylan Herts, and Juan Pablo. "Analysis and detection of modern spam techniques on social networking sites." In 2012 Third International Conference on Services in Emerging Markets, pp. 147-152. IEEE, 2012.*

[6] *Beck, Kristofer. "Analyzing tweets to identify malicious messages." In 2011 IEEE International Conference on Electro/Information Technology, pp. 1-5. IEEE, 2011.*

[7] *Lin, Po-Ching, and Po-Min Huang. "A study of effective features for detecting long-surviving Twitter spam accounts." In 2013 15th International Conference on Advanced Communications Technology (ICACT), pp. 841-846. IEEE, 2013.*